Avoiding the Crowds: Understanding Tube Station Congestion Patterns from Trip Data

Irina Ceapa
Dept. of Computer Science
University College London
London WC1E 6BT, UK
i.ceapa@cs.ucl.ac.uk

Chris Smith
Dept. of Computer Science
University College London
London WC1E 6BT, UK
c.smith@cs.ucl.ac.uk

Licia Capra
Dept. of Computer Science
University College London
London WC1E 6BT, UK
l.capra@cs.ucl.ac.uk

ABSTRACT

For people travelling using public transport, overcrowding is one of the major causes of discomfort. However, most Advanced Traveller Information Systems (ATIS) do not take crowdedness into account, suggesting routes either based on number of interchanges or overall travel time, regardless of how comfortable (in terms of crowdedness) the trip might be. Identifying times when public transport is overcrowded could help travellers change their travel patterns, by either travelling slightly earlier or later, or by travelling from/to a different but geographically close station. In this paper, we illustrate how historical automated fare collection systems data can be mined in order to reveal station crowding patterns. In particular, we study one such dataset of travel history on the London underground (known colloquially as the "Tube"). Our spatio-temporal analysis demonstrates that crowdedness is a highly regular phenomenon during the working week, with large spikes occurring in short time intervals. We then illustrate how crowding levels can be accurately predicted, even with simple techniques based on historic averages. These results demonstrate that information regarding crowding levels can be incorporated within ATIS, so as to provide travellers with more personalised travel plans.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Algorithms, Experimentation

Keywords

Mobility, Public Transport, Predictions

1. INTRODUCTION

With over 75% of the world's population expected to be living in cities by 2050, supporting citizens mobility within the urban environment is a priority for municipalities worldwide. In order to better manage mobility, public multimodal transit systems, coupled with integrated fare management, are being deployed in large cities, including London, UK. However, these transit systems are not able to absorb increasing loads and, especially at peak commuting hours when the system has to cope with a temporary surge in demand, overcrowding creates incredibly high levels of discomfort[1]. As a result, car ownership continues to grow worldwide, despite financial disincentives like costly congestion charges.

How can we overcome the problem of overcrowding in urban public transport systems? Transport operators often attempt to discourage peak-time travel by means of fare differentiation. For example, Transport for London (TfL) uses two price bands for morning-peak and post-morning-peak travel (before and after 9:30AM respectively), with the former being up to 50% more expensive than the latter. However, this has had no observable effect on travellers' journeys [9], with the tube reaching its peak load at around 8:15AM every weekday. This suggests that peoples' behaviour cannot be changed with imposed fare policies that do not take into consideration external constraints (e.g., having to be at work by 9:00AM in the morning). Instead we propose a different method for encouraging travellers to diversify their habits, by providing information about the likely crowding levels.

We analyse underground station usage in London, starting from Automated Fare Collection (AFC) systems data. AFC systems forgo traditional fare media, such as paper tickets or magnetic strip cards, in favour of alternatives such as RFID-based smart cards (e.g., London's Oyster Card, Seattle's Orca Card) or near-field communication on mobile phones (e.g., the Tokyo Metro System). These new payment systems create a digital record every time a trip is made, recording, for example, origin, destination, and travel time of every journey made. By analysing this data we discover that during the working week, crowdedness is a highly regular phenomenon, most probably as a direct result of the home-work commutes that follow rather fixed schedules. Even more interestingly, we show that spikes of crowdedness are concentrated in very short time periods, leaving the transport network under-utilised before and after such spikes. These findings suggest that, if made aware of such crowdedness patterns, travellers could opt to depart slightly earlier or later, thus avoiding congestion peaks while still

UrbComp '12, August 12, 2012. Beijing, China

[1] http://www.guardian.co.uk/money/2011/dec/30/worst-train-reading-london-paddington

making it to work/home on time. We then build a number of crowdedness predictors and compare their accuracy while varying crowdedness thresholds, the size of the prediction window, and amount of data used for training such predictors. In all cases, we find accuracy is very high, even when using very simple techniques.

The remainder of this paper is structured as follows: Section 2 offers background information about the London Underground system and the AFC dataset we have used. In Section 3 we present the results of a spatio-temporal analysis of such data, illustrating the high regularity and short-lived spikes with which crowdedness manifests itself at different stations. Based on this finding, we move from analysis to prediction, and in Section 4 we demonstrate how station crowding levels can be accurately predicted from historical data, even with simple techniques based on historic aggregates. Section 5 positions this work with respect to similar ongoing efforts by the research community. Finally, in Section 6 we present our future work, where we plan to conduct choice experiments with a large population sample, to assess whether travellers' behaviour can be nudged simply by offering them accurate information about the crowdedness levels of the transport network.

2. THE OYSTER CARD DATA

The London Underground network is made up of 11 lines totalling 402 kilometres of track, serving near 270 stations. The transport network is separated into 9 fare zones, with Zone 1 encompassing central London and higher numbers representing regions further away from the centre of London, up to Zone 9, which contains a handful of outlier stations. The zoning system forms part of the fee structure for all rail travel in London, as well as approximating geographical distance from the centre of London. It is estimated that approximately 1,107 million passengers are carried by the tube each year. TfL introduced the Oyster Card in 2003 which now accounts for over 80% of all trips made within London's public transport system (as opposed to traditional paper tickets). Detailed information about each trip is captured when an Oyster card is used to enter or exit the tube network, producing an extensive source of data. For this study, we used a dataset containing all trips made with an Oyster card in the 31 days of March 2010. Each trip in our dataset is recorded as a tuple in the form:

$$\langle u, (o, d), t_o, t_d \rangle$$

Each tuple contains the unique (anonymous) user id (u), the trip (with origin o and destination d stations), the time stamp (to minute precision) t_o when u entered the origin station and the time t_d when u exited from the destination station. Over 70 million trips were recorded in March 2010. After cleaning the raw dataset (e.g., eliminating journeys with an end time that was before the start time and trips where the origin was the same as the destination), approximately 5 million trips ($\approx 8\%$) were removed, leaving us with over 64 million trips in total.

In order to analyse crowdedness at stations, we transformed this data from a per-user basis to a per-station basis as follow: each tuple $\langle u, (o, d), t_o, t_d \rangle$ was split in two, one recording station and time of origin $\langle o, t_o \rangle$ (*touch-ins*), and one recording station and time of exit from destination $\langle d, t_d \rangle$ (*touch-outs*). For each of the 270 stations, raw numbers were aggregated on 2-minute intervals, meaning that, for each day, 720 observations (i.e., the number of touch-ins or touch-outs in the 2-minute interval) were recorded.

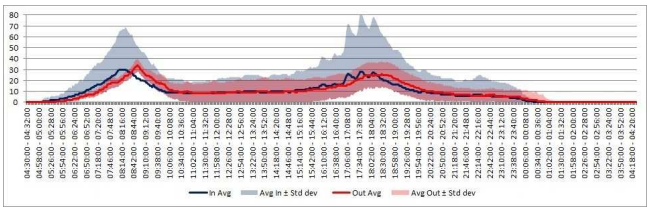

Figure 1: System-wide Weekday View

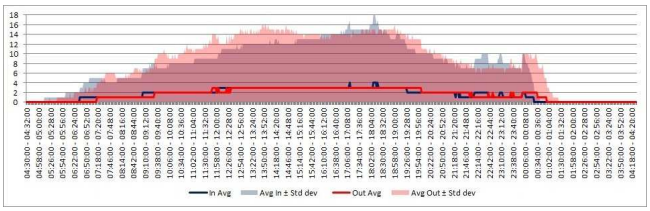

Figure 2: System-wide Weekend View

3. CROWDEDNESS ANALYSIS

Having pre-processed the data as described above, we next analyse the aggregate, system-wide usage patterns across all stations, to give a broad perspective of the usage of the system. We then demonstrate the variation in activity patterns exhibited by different by different stations by focusing on a select few. Based on these insights, we then move onto a spatio-temporal analyses of the system, where we use agglomerative hierarchical clustering (a form of unsupervised learning) to classify different patterns which emerge at different stations.

3.1 System-wide Temporal Analysis

To begin with, we look at system-wide aggregate data, only distinguishing between weekdays (Figure 1) and weekends (Figure 2). As expected, the weekday aggregated system behaviour is dominated by the two-pronged commuter pattern, with peaks in the morning at around 8:15AM, and in the afternoon at around 5:30PM, with approximately 35 touch-ins or touch-outs for every 2-minute interval. There are two interesting observations to make about the evening commute period. Firstly, the curve is not as steep as it was for the morning commute. This might be explained by the fact that, in the morning, people have to arrive at work by a certain time, whereas they are not under such an obligation on the return journey in the evening. Secondly, the touch-in curve during the evening peak time displays three smaller prongs, with peaks at roughly every 30 minutes, staring from about 5:10PM. This can also be motivated by people's work behaviour: as the standard working day is from 9:00AM to 5:00PM, some people leave work just after 5:00PM, while others leave later, but organise their schedules around specific time intervals, such as 30 minutes. Note also the large standard deviation (shaded area in the plot), which indicates that the data hides additional information or patterns which we explore next.

The weekend aggregate behaviour, illustrated in Figure 2, is not only much less regular with very high standard

deviation, suggesting that crowding levels will not be as easy to predict, but also much less intense (y axis), which suggests that overcrowding is generally not a problem at weekends. For these reasons we thus disregard weekend data, and focus on weekdays only.

3.2 Spatio-Temporal Analysis

The system-wide analysis suggests that the weekday activity is very regular and therefore predictable. However, wide standard deviations suggest that a closer inspection may reveal hidden patterns in the data. For example, we would expect to see a difference between stations in residential areas versus stations in areas that mainly contain office buildings, since the flow of traffic will be in opposite directions during commuting periods. Furthermore, specific patterns should be observed in stations close to sporting locations, or stations close to cultural and entertainment locations, such as theatres. To investigate this we manually pick three stations representative of the different station usages mentioned: a station in a residential area, a station in a busy financial area, and a transport hub station, linking to National Rail.

Residential station – Finchley Central. Finchley Central serves a residential area and we expect this to be reflected in the commute pattern. Weekday activity is illustrated in Figure 3. The morning commute is dominated by touch-ins – people leaving the station and going to work in another area of the city. The maximum number of touch-ins is 60, at around 8:15AM. An important observation is the fact that the standard deviations for this period are relatively low, suggesting regular behaviour. During the evening commute, the situation is reversed, with touch-outs dominating the station usage pattern. As observed in the system-wide analysis, the evening commute is much less steeper than the morning commute. The standard deviations are also much wider, suggesting that the travel patterns in the evening are less regular.

Business station – Canary Wharf. Canary Wharf serves one of the two main financial centers of London, with over 93,000 people work in the area[2], which is mainly served by the Canary Wharf tube station. We expect the travel patterns to be the reverse of those encountered in stations serving residential areas. The day view for weekdays is illustrated in Figure 4. The morning activity consists almost entirely of touch-outs, with people arriving in the area. It increases steadily until it peaks at around 9:00AM, with 500 people leaving the station every 2 minutes. The standard deviation for the morning commute is quite high, reaching almost 100 touch-ins at the peak of the morning commute. As expected, the evening commute exhibits the opposite behaviour, with almost no touch-outs and an extremely regular (indicated by the very small standard deviation) three-pronged spike of touch-ins. The three prongs are spaced approximately 30 minutes apart.

Transport hub station – Waterloo. As our third and final case study, we consider Waterloo, a national railway terminus and busy tube interchange in central London. This is the busiest tube station in the entire TfL network, achieving more than 80 million touch-ins and touch-outs every year. Unlike the previously discussed station profiles, the weekday view, illustrated in Figure 5, shows much smaller

[2]`http://www.canarywharf.co.uk/aboutus/The-Estate/General-Information/`, retrieved 20 March 2012

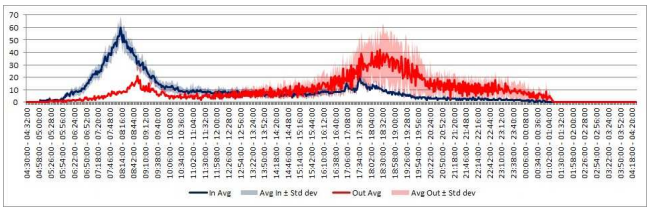

Figure 3: Finchley Central Weekday View

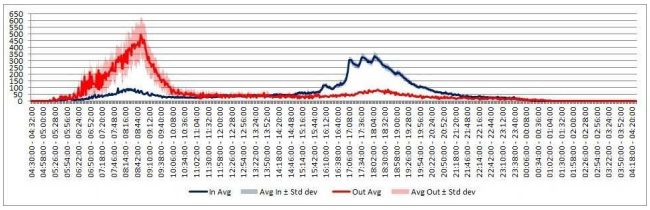

Figure 4: Canary Wharf Weekday View

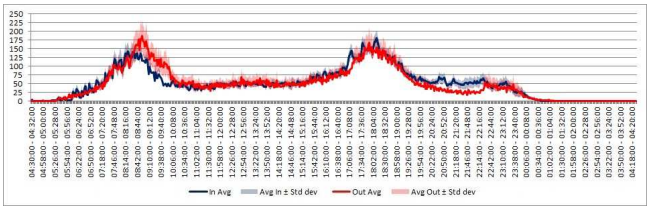

Figure 5: Waterloo Weekday View

differences between touch-ins and touch-outs. Indeed, the peaks for the touch-ins and touch-outs are quite similar in magnitude and only slightly out of phase.

The empirical analysis above supports the supposition that the dominant commute pattern hides individual station usage patterns. We thus proceeded with a more systematic approach, where we used an agglomerative hierarchical clustering technique to automatically group stations based on emerging usage patterns [18]. More precisely, to measure similarity between station usage patterns, we used a technique called dynamic time warping [19], a well-known algorithm used for computing differences between time series. The main strength of the algorithm is that it is extremely efficient as a time-series similarity measure which minimizes the effects of shifting and distortion in time by allowing elastic transformation of time series in order to detect similar shapes with different phases. In our analysis, we used an approximation to Dynamic Time Warping called FastDTW[3], that has linear time and space complexity, proved both theoretically and empirically [17]. The same study also proves that it shows a large improvement in accuracy over two other existing approximate DTW algorithms, Sakoe-Chuba Bands and Data Abstraction.

Using this similarity metric we constructed a hierarchy of clusters using an agglomerative approach. The algorithm works by starting with as many clusters as stations, then iteratively merging the two most similar clusters until a specified halting condition is met. The similarity between clusters

[3]A Java implementation is provided under an open-source MIT licence at `http://code.google.com/p/fastdtw/`

	C1	C2	C3	C4	C5	C6
Num. of stations	1	23	8	27	3	198
Intra-cluster distance	0.000	0.198	0.008	0.003	0.134	0.011

Table 1: Clusters Information

	C1	C2	C3	C4	C5	C6
C1	0.0	2.316	1.167	1.992	1.205	2.115
C2		0.0	1.745	1.296	1.756	1.206
C3			0.0	1.495	1.395	1.531
C4				0.0	1.546	1.132
C5					0.0	0.550
C6						0.0

Table 2: Inter-Cluster Distances

is defined as the *average linkage*:

$$D_{AB} = \frac{1}{n_A n_B} \sum_{a \in A} \sum_{b \in B} FastDTW(a, b)$$

where clusters A and B have n_A and n_B members (stations) respectively.

The input to the clustering algorithm consists of a vector for each station, containing the difference between touch-ins and touch-outs at each 2-minute time interval. Thus, we characterise each station as either a sink (touch-outs > touch-ins, i.e., more people arrive to this station than leave from this station) or a source (touch-outs < touch-ins). As the size of stations varies significantly, in order to compare congestion levels, we also normalised the data, by dividing the difference between touch-outs and touch-ins by the largest of maximum number of touch-ins and touch-outs. The agglomerative clustering algorithm was then terminated when 6 clusters were produced; this number was found by visual inspection to provide the most informative array of station activity patterns.

A summary of the six resulting clusters is given in Table 1. The majority of the stations were included in cluster 6, while clusters 2 and 4 have between 20 and 30 stations. Clusters 3 and 5 are very small (8 and 3 stations, respectively), while cluster 1 only has 1 station (Wood Lane). The *intra*-cluster distances are very small, suggesting good cluster compactness. Table 2 shows the *inter*-cluster distance. Compared to the intra-cluster distances the inter-cluster distances are large, suggesting a good cluster separation.

The average day view for each of the six clusters is displayed in Figure 6. Sink-like behaviour will result in positive values, while source-like behaviour will display negative values. Cluster 1 is the most visually distinct of the six clusters, with high values and a morning peak which occurs significantly later than the dominating commute pattern. It acts as a powerful sink during the morning, with a peak at around 10:00AM. This behaviour is slowly attenuated until around 5:00PM, when the station becomes a source, although relatively small absolute values suggest that the average number of touch-outs is only slightly larger than the touch-ins. Cluster 2 exhibits opposite behaviour, acting as a source during the morning commute and as a sink during the evening one. Its peaks occur between 7:30AM and 8:00AM, and between 6:30PM and 7:00PM, respectively. Clusters 4 and 6 are very similar to cluster 2, and differ only in the amplitude of the absolute values and in the time when they reach the morning peak (between 8:00AM and 8:30AM). Clusters 3 and 5 are different from all other clusters in the sense that they change their behaviour during the morning commute and again during the evening one. Both start out as sources, until around 8:30AM, when they become sinks before stabilising at a neutral behaviour around mid-day. During the evening, they both return to source behaviour, although cluster 5 turns into a sink between 6:30PM and 9:30PM.

4. CROWDEDNESS PREDICTION

In the previous section we found that tube stations in London are dominated by the commute usage pattern (during weekdays), but with a significant number of stations exhibiting more distinct usage patterns. In all cases, we note that the usage patterns are highly regular during weekdays, so we expect to be able to predict usage levels and, consequently, overcrowding. In this section, we formulate the crowdedness prediction problem as a classification problem (Section 4.1), before reporting the results of an extensive evaluation that compares accuracy as obtained with different classification techniques (Section 4.2).

4.1 Methodology

The Dataset. Recall that we are working with weekdays only. In our dataset, we have 23 days of data (once we exclude weekends), which were divided into two sets: a training set (first 18 days), used to calibrate the parameters of the predictors, and a testing set (last 5 days), used to evaluate the performance of the predictors. While in the analysis section we worked with 2-minute intervals, for prediction we decided to work with 10-minute intervals. This is because an important follow-up of this study is to see whether information on congestion could nudge people to adapt their travel patterns, perhaps by leaving earlier or later. Using too fine-grained intervals (such as 2 minutes) would be too short for people to be able to appreciate the difference and adapt. On the other hand, too coarse-grained intervals (such as 30 minutes) would be impractical and have too much of an impact on peoples' schedules, resulting in people choosing not to adapt. We decided on using 10-minute intervals because they seem the most natural choice to nudge people into changing the time at which they travel. The training and testing data files for each station contain 144 observations per day.

The λ Threshold. To predict whether a station is crowded or not at a given point in time, we first need to define what it means for a station to be crowded. In the absence of official station capacities and congestion thresholds([14]) we define a proxy measure of crowding level and experiment with varying congestion thresholds. We use as a measure of crowding level the proportion of touch-ins (or touch-outs) at the station relative to the maximum number observed in the data. Thus, the maximum crowdedness of 1 indicates the station is at its peak level of crowdedness, and conversely, 0 indicates no touch-ins (or touch-outs) within the measurement interval. Identifying an appropriate congestion threshold, λ, is itself an interesting research question pertaining to individual travellers' preferences. Indeed, crowding tolerance

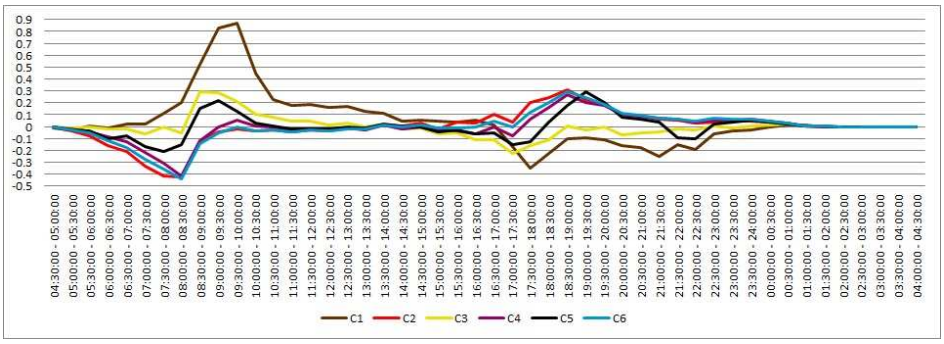

Figure 6: Average Day Views for the Six Clusters

varies greatly among individuals, and even for the same individual in different circumstances (imagine having to get to work in time for a meeting versus planning a trip to go for a walk in Hyde Park). Therefore, we decided to run and evaluate the predictors with different λ values. As we have until now, we will treat touch-ins and touch-outs separately (i.e., crowdedness entering vs leaving a station). The total number of overcrowded intervals for different values of λ is reported in Table 3. Based on these results, we decided to use 3 different values for λ: 0.5, 0.6 and 0.8. As we shall see, as the value of λ increases, and hence the classification problem becomes stricter, the accuracy of the results of our predictors will gradually decline.

λ	Overcrowded (in)	overcrowded (out)
0.4	20.92	25.76
0.5	15.05	17.67
0.6	10.78	12.50
0.7	7.51	8.63
0.8	4.96	5.43

Table 3: The effect of λ on the number of overcrowded intervals

Metrics. We measure results of our classification problem in terms of true positives, true negatives, false positives and false negatives. If both the predicted crowding level and the observed crowding level are greater than or equal to our crowding threshold, λ, the result is a true positive tp. If the predicted crowding level is greater than or equal to λ but the real level is not, the result is a false positive fp. Classifications for negative results (true negative tn and false negative fn) are determined in a similar way. The most undesirable result would be for us to incorrectly predict that an interval is not overcrowded. This would mean that travellers relying on our predictions to avoid congestion would be faced with an overcrowded station. In other words, our predictors should be evaluated primarily with how accurately they avoid false negatives. For this reason, we report results in this paper in terms of *sensitivity*, that is:

$$Sensitivity = \frac{tp}{tp + fn}$$

Sensitivity (also known as *true positive rate* or *recall*) is the proportion of correctly identified positive results from all positive results. For our specific problem, it indicates how often we are right that a station is overcrowded. For a more complete evaluation across other metrics, including precision ($tp/(tp + fp)$), accuracy ($((tp + tn)/(tp + fp + tn + fn))$), specificity ($(tn/(tn + fp))$) and F-measure ($2 \cdot (Precision \cdot Sensitivity)/(Precision + Sensitivity)$), the interested reader may refer to [1].

Techniques. We ran three different prediction algorithms, each taking input t, the current time interval, and PW, the prediction window, which varies from 10 (i.e., predict crowding level in the next 10 minute interval) to 120 minutes (i.e., predict crowding level for a 10 minute interval that starts 110 minutes from t). Using the following notation:

- $Train[t]$ – the average crowding level at time interval t in the training set,
- $Test[t]$ – the observed crowding level at time interval t in the test set,
- $\overline{Train[t_1 - t_2]}$ – the average of the crowding levels in the training set during the time intervals from t_1 to t_2, inclusive,

we can formally define our three predictors as follows:

Historic Value - this predictor reports, for all time intervals and for all values of the prediction window, the corresponding value at the interval $t + PW$ in the training set. Such a baseline predictor is expected to perform well if and only if crowdedness is extremely regular (e.g., crowdedness on a Friday at 9:00AM-9:10AM is the same as the average of the recorded crowdedness levels on all weekday slots 9:00AM-9:10AM in the training set).

$$HistoricValue(t, PW) = Train[t + PW]$$

Historic Mean - this predictor also takes advantage of the history available in the training set. However, for a given time interval t and prediction window PW, it reports the average of all values in the training set between t and $t + PW$. This predictor is expected to perform well if crowdedness levels are regular (reliance on historic averages) but temporally shifted within a time window (e.g., crowdedness on a Friday at 9:00AM-9:10AM is the average of the recorded crowdedness levels on all weekday slots from 8:40AM to 9:10AM in the training set).

$$HistoricMean(t, PW) = \overline{Train[t - (t + PW)]}$$

Historic Trend - this technique attempts to improve the Historic Mean predictor by taking into consideration crowdedness level as currently recorded. This means that it can

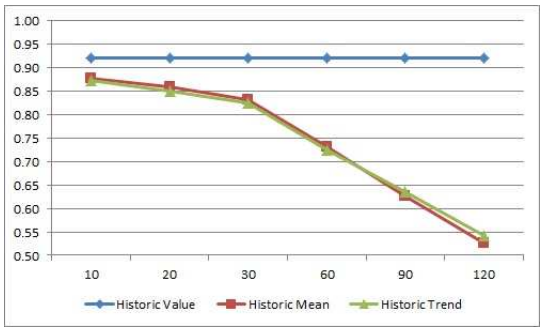

Figure 7: Sensitivity for $\lambda = 0.5$ (touch-ins)

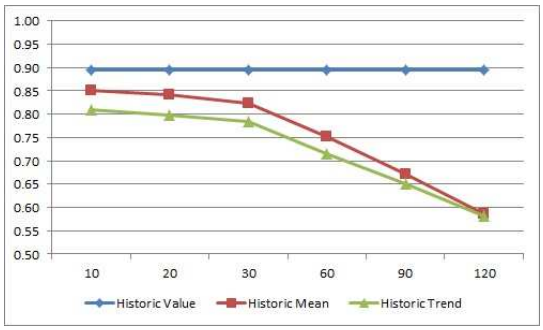

Figure 8: Sensitivity for $\lambda = 0.5$ (touch-outs)

take into consideration anomalies, such as outliers, in the training set that could influence the result. To achieve this, the Historic Trend implementation replaces the value at the current interval in the training set with the corresponding value in the testing set.

$$HistoricTrend(t, PW) = \overline{Train[t-(t+PW)]} - Train[t] + Test[t]$$

We also experimented with two other predictors, one based on linear regression with ordinary least squares to estimate parameters, and one based on Kalman Filters. Neither techniques offered improvements over the three techniques above, so we leave them out of this paper in the interest of space. A full report of the findings is available at [1].

4.2 Results

Fixing $\lambda = 0.5$. We begin our evaluation by presenting the results obtained by the three predictors with a congestion threshold of $\lambda = 0.5$. Results in Figures 7 and 8 are aggregated and averaged over all stations in the system, separately for touch-ins and touch-outs. As shown, all predictors perform well when considering short prediction windows. As the prediction window increases from 30 minutes up to 2 hours away, the performance of Historic Mean and Historic Trend worsens, while leaving the performance of Historic Value unscathed; this confirms our hypothesis that crowdedness is highly regular and highly spiked too.

Impact of Varying λ. As previously discussed, the reasons for choosing a particular value for the crowdedness threshold, λ could be explored by an entire user study in its own right, focusing on what crowdedness means for different people or even for the same individual in different

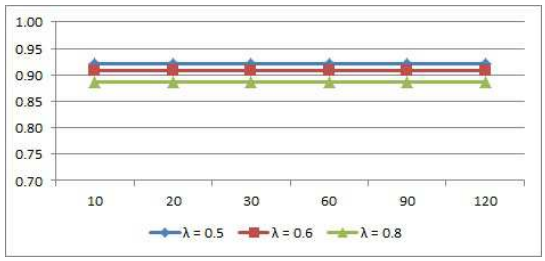

Figure 9: Effect of λ on Sensitivity of Historic Value (touch-ins)

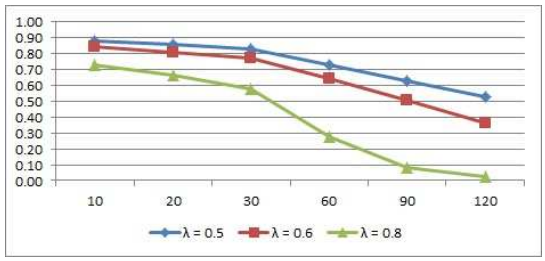

Figure 10: Effect of λ on Sensitivity of Historic Mean (touch-ins)

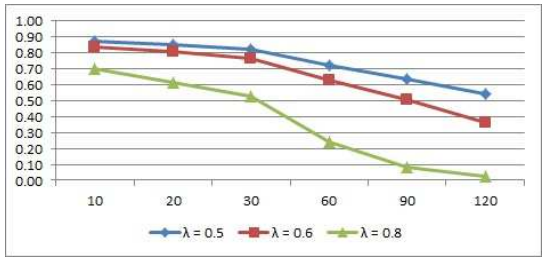

Figure 11: Effect of λ on Sensitivity of Historic Trend (touch-ins)

circumstances. We intend to do so in our future work. In this paper, we limit ourselves to studying the impact of the crowdedness threshold on the performance of the prediction techniques. In the interest of space, we show results for touch-ins only (results for touch-outs are only slightly worse and can be found in [1]).

Figures 9, 10 and 11 illustrate the effect of the congestion threshold values on the performance of the Historic Value, Historic Mean and Historic Trend predictors respectively. As the crowdedness threshold increases, performance decreases (though only marginally so for Historic Value). This result should partly be interpreted in light of the data we have: recall from Table 3 that, as λ increases, the number of crowded time intervals (true positives) significantly decreases. As such, mis-classifying even a single interval results in a quick performance decrease (as measured by the sensitivity metric).

Impact of Training Data. As all our predictors rely on average historic data, an interesting question to answer is how much history is required in order to reach high sensitivity. In other words, what would be the effect of changing the ratio of the training and testing sets. We again evaluate the performance of our three predictors, that is, Historic

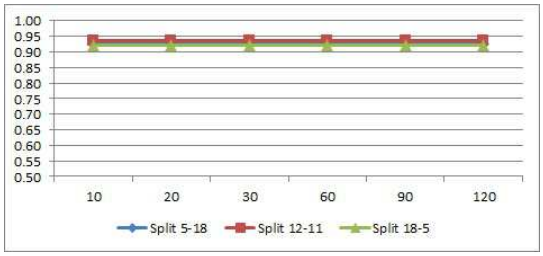

Figure 12: Effect of History Data on Sensitivity for Historic Value (touch-ins)

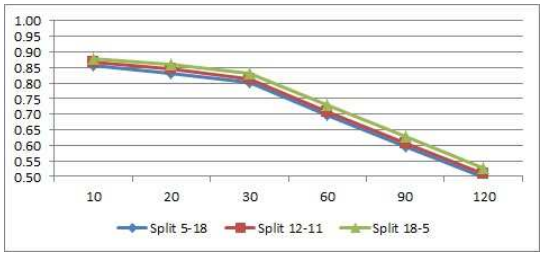

Figure 13: Effect of History Data on Sensitivity for Historic Mean (touch-ins)

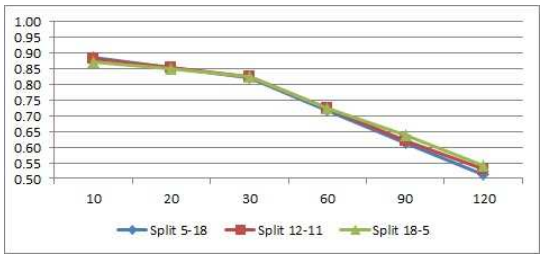

Figure 14: Effect of History Data on Sensitivity for Historic Trend (touch-ins)

Value, Historic Mean and Historic Trend, and results are shown in Figures 12, 13, and 14 respectively. Rather than using 18 days of training and 5 days of testing, we now test two splits: 12 days training - 11 days testing ($12-11$), and 5 days training - 18 days of testing ($5-18$). Once again, in the interest of space, we show results for touch-ins only.

No noticeable difference can be see for Historic Value, with consistently good sensitivity shown throughout. In the case of the Historic Mean predictor, the higher the amount of training data, the better the results, but once again the differences are not very pronounced (no more than 5% improvement). Finally, the Historic Trend predictor also shows better performance with more historical data, but the difference between the 3 splits is very small (about 1-2% on average).

Having only one month worth of data, one has to be cautious with the conclusions being made in terms of prediction accuracy. However, the above analysis would suggest that even short periods of training data (e.g., 2 weeks) are sufficient for the above predictors to learn station usage patterns, as these tend to be very regular. We can thus accurately predict crowdedness at stations, and consequently build more sophisticated journey planner tools that take this factor into account. It is worth noting that, in this work, we are interested in measuring and predicting crowdedness due to seasonal movements, and not exceptional situations due to, for example, unforeseen faults in the system (in other words, such anomalies are smoothed out by our predictors). Anomaly detection would be very useful to offer real-time information during journey execution, so that travellers can adapt on the go. This is however a complementary area of work which we do not investigate in this paper.

5. RELATED WORK

The increasingly wide availability of AFC data has lead to an explosion of research primarily focused on how such data can be used to evaluate and study the performance of the transportation system itself. For example, through demand modelling [3], service reliability measurements [2], average travel time estimation [2], and station transfer analysis [7]. A complimentary line of work has been looking at what the AFC data reveals, not about the transportation system, but about individual traveller behaviour instead: for example, by offering personalised travel time estimations [11], or by recommending what ticket type to purchase [10].

In this paper we start to combine the two: that is, we use AFC data to measure and estimate crowdedness levels of various stations in the transport network (system focus), but with the aim to feed this information back to the traveller, in terms of personalised classifications (station [not] crowded) based on individual tolerances to crowdedness (the λ value). In so doing, this paper adds to the growing body of work on smart cities and urban informatics [4], which is the study of human behaviours and urban infrastructures made possible by the increasingly digitised and networked city. For example, Gonzalez et al., 2008 [6] and Ratti et al., 2008 [15] use mobile phone-based location data to study human mobility patterns; Kostakos et al., 2006 [8] and Sadabadi et al., 2010 [16] rely on distributed Bluetooth receivers to track and predict travel speeds based on the Bluetooth MAC identifiers of passing devices. Most of the cited work, however, continues to focus on aggregate analysis only, rather than attempting to uncover opportunities for personalisation services.

One limitation of our work is that we study 'station' crowdedness only, while it would be useful for travellers to have also information about 'train' crowdedness. Oyster card data prevents us from building trajectory or sub-route-based models (e.g., [12], [5], [20]) since the actual route that a user undertakes between any origin and destination is unknown to us; in many cases, there are a wide variety of candidate routes. Implementing heuristics to derive route choices (for example, minimising the number of interchanges or minimising the hop-count on the tube graph) does not resolve cases where two routes seem equal on the applied heuristic (e.g., they both have one interchange) or when the heuristic derives results where travel time may increase (e.g., in cases where changing line would have reduced travel time). An area of future work will be to incorporate additional sensory information, such as images captured by CCTV cameras installed at station platforms, so to accurately quantify and subsequently predict train crowdedness. So far, CCTV cameras have only been used to detect (and not predict) 'platform' congestions [13]; this is useful for staff members to decide, for example, when to temporarily close stations in entry/exit to alleviate overcrowding.

6. CONCLUSION AND FUTURE WORK

In this paper, we have analysed anonymised AFC data collected for the tube network in London, UK, and have shown that crowding levels at stations is highly regular, and can thus be accurately predicted using simple predictors based on historic data averages. Not only did we show that crowdedness is a highly regular (and predictable) phenomenon during the working week, we also highlighted that big spikes concentrates in rather short time periods. The most important question for us now is what would people do with the information uncovered by this study? Would they change their behaviour, either by adjusting their travel times (by a small interval) to avoid the congestion peak, or by travelling to a different, yet close (geographically or on the same line) tube station? A user study is now required in order to determine what impact, if any, congestion information has on travel patterns. As part of this study, we also need to determine what the congestion threshold λ means for different people and in different situations, so that journey planner tools can leverage this information in computing personalised routes.

Acknowledgements. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7- SST-2008-RTD-1) under Grant Agreement n. 234239.

7. REFERENCES

[1] I. Ceapa. Predicting tube station congestion. Master's thesis, University College London, April 2012.

[2] J. Chan. Rail Transit OD Matrix Estimation and Journey Time Reliability Metrics Using Automated Fare Data. Master's thesis, MIT. June 2007.

[3] K. Chu and R. Chapleau. Enriching Archived Smart Card Transaction Data for Transit Demand Modeling. *Journal of the Transportation Research Board*, (2063), 2008.

[4] M. Foth. *Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City*. 2009.

[5] J. Froehlich and J. Krumm. Route Prediction From Trip Observations. In *Intelligent Vehicle Initiative, SAE World Congress*, Detroit, Michigan, 2008.

[6] M. Gonzalez, C. Hidalgo, and A.-L. Barabasi. Understanding Individual Human Mobility Patterns. *Nature*, 453(7196):779–782, 2008.

[7] W. Jang. Travel Time and Transfer Analysis Using Transit Smart Card Data. *Journal of the Transportation Research Board*, (3859), 2010.

[8] V. Kostakos, T. Kindberg, and et al. Instrumenting the City: Developing Methods for Observing and Understanding the Digital Cityscape. In *In Ubicomp*. 2006.

[9] N. Lathia and L. Capra. How smart is your smartcard?: measuring travel behaviours, perceptions, and incentives. In *Ubicomp*, 2011.

[10] N. Lathia and L. Capra. Mining mobility data to minimise travellers' spending on public transport. In *17th ACM SIGKDD*, 2011.

[11] N. Lathia, J. Froehlich, and L. Capra. Mining public transport usage for personalised intelligent transport systems. In *ICDM*, 2010.

[12] H. V. Lint. Empirical Evaluation of New Robust Travel Time Estimation Algorithms. In *89th Annual Transport Research Board*, January 2010.

[13] B. Lo and S. Velastin. Automatic congestion detection system for underground platforms. In *Intl. Symposium on Intelligent Multimedia, Video and Speech Processing,*, May 2001.

[14] Freedom of Information request to TfL. Station capacity and safety policy. `http://www.whatdotheyknow.com/request/station_capacity_and_safety_poli`. answered 17 January 2012.

[15] C. Ratti, R. M. Pulselli, S. Williams, and D. Frenchman. Mobile Landscapes: Using Location Data From Cell Phones for Urban Analysis. *Environment and Planning B: Planning and Design*, 33(5):727–748, 2006.

[16] K. Sadabadi, M. Hamedia, and A. Haghani. Evaulting Moving Average Techniques in Short-Term Travel Time Prediction Using an AVI Data Set. In *89th Transportation Research Board Annual Meeting*.

[17] S. Salvador and P. Chan. Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis*, 11(5):561–580, Oct. 2007.

[18] J. L. Schmidhammer. Agglomerative hierarchical clustering methods. `http://bus.utk.edu/stat/stat579`.

[19] P. Senin. Dynamic time warping algorithm review. *Information and Computer Science Department, University of Hawaii at Manoa, Honolulu, USA*, Dec. 2008.

[20] Y. Zheng and X. Zhou. Computing with Spatial Trajectories. Springer 2011.